Microfilariae Classification Using Multiple Classifiers for Color and Shape Features

AL-Tam, Faroq; Dos Anjos, António; Pion, Sébastien; Boussinesq, Michel; Shahbazkia, Hamid Reza

Published in:
Open Engineering

DOI:
10.1515/eng-2016-0079

Publication date:
2016

Citation for published version (APA):
AL-Tam, F., Dos Anjos, A., Pion, S., Boussinesq, M., & Shahbazkia, H. R. (2016). Microfilariae Classification Using Multiple Classifiers for Color and Shape Features. *Open Engineering*, *6*(1), 560-565. https://doi.org/10.1515/eng-2016-0079

ICEUBI 2015* Open Access

Faroq AL-Tam*, António dos Anjos, Sébastien Pion, Michel Boussinesq, and Hamid Reza Shahbazkia

Microfilariae Classification Using Multiple Classifiers for Color and Shape Features

DOI 10.1515/eng-2016-0079
Received Apr. 08, 2016; accepted Sep. 06, 2016

Abstract: This paper presents a multi-classifier approach for classifying microfilariae in 2-D images. A shape descriptor based on the quench function is described. This descriptor is represented as a feature vector that encodes the shape information. The color feature vector is calculated as a histogram. Two classifiers were used to train both color and shape feature vectors, one for each vector. The posterior probabilities calculated from the scores of each classifier are then used to calculate the final classification decision. The experimental results show that, although the proposed approach is simple, it is efficient when compared to various approaches.

Keywords: Loa loa; loiasis; microfilariae; microscopy imaging; illumination correction; multi-classifier

1 Introduction

Digital microscopy image analysis is a research field that arouses an increasing interest in the biomedical world. The amount of data obtained from biological experiments justifies the need for automatic processing. Common uses for this kind of analysis is the diagnosis of diseases and/or drug design. In this context, the manual analysis of microscopic particles is very time-consuming and error-prone (if not impossible). This work addresses the automatic analysis of a type of microscopic images used in study and diagnosis of a common parasitic disease, loaisis, which is a major health problem in Central Africa.

Loiasis is caused by a filarial nematode (round worm) called *Loa loa* [1]. This disease affects over 12 million people in more than 11 African countries. The main vectors of *L. loa* are two species of tabanids: *Chrysops silacea* and *C. dimidiata*. The most frequent manifestations of loiasis are transient lymphoedemas (known as "Calabar swellings") usually occurring on the upper limbs, and the migration of the adult stage of the parasite under the conjunctiva of the eye (loiasis is also known as the "African eye-worm"). Adult worms live under the skin or between muscles, and females produce millions of embryos, called microfilariae, that invade the blood circulation from where they can be picked up by the vector to maintain the parasite cycle. In addition to being a health issue in itself, loiasis is also a subject of interest because individuals harbouring high microfilarial densities, can develop serious adverse events (SAEs) after treatment with the drug ivermectin [2, 3]. Ivermectin is widely used to control two other filarial diseases which constitute public health problems in Africa: onchocerciasis known as "river blindness" and lymphatic filariasis known as "elephantiasis".

Although the clinical manifestations mentioned above, are pathognomonic, the standard method to diagnose *L. loa* infection is by microscopic examination of the blood smears prepared with blood drawn from the host. This kind of diagnosis involves a process of staining and optional concentration to demonstrate the presence of microfilariae in the blood. In most cases, giemsa-staining is used, when possible, and the Knott's technique using formalin can be applied for concentration [4]. However, the microscopy slides may contain debris and unwanted particles (dust, etc.). Moreover, images are often affected by a biased illumination field when improper microscope settings are used. The objective of this work is to develop an approach to identify microfilariae in these kinds of

***Corresponding Author: Faroq AL-Tam:** DCT, ISMAT, 8500-508 Portimão, and University Algarve, 8005-139 Faro, Portugal; Email: ftam@ualg.pt
António dos Anjos: Department of Computer Science, Aberystwyth University, Aberystwyth, SY23 3DB, United Kingdom; Email: ara11@aber.ac.uk
Hamid Reza Shahbazkia: FCT, University Algarve, 8005-139 Faro, Portugal; Email: hshah@ualg.pt
Sébastien Pion, Michel Boussinesq: UMI 233, Institute of Research and Development and University of Montpellier, Montpellier, France; Email: {michel.boussinesq,sebastien.pion}@ird.fr

* International Conference on Engineering 2015 – 2–4 Dec 2015 – University of Beira Interior – Covilhã, Portugal

images. Such a technique would be most useful to identify rapidly in routine those individuals who have high microfilarial densities, and thus have a risk to develop a post-ivermectin SAE. Using a test and treat strategy, at risk individuals (who represent less than 5% of the total population) could be excluded from mass treatment with ivermectin, whereas the rest of the population could be treated safely with the drug.

1.1 Related Work

Microfilariae recognition in 2D images is scarce in literature. Nevertheless, two different groups of methods can be reviewed here. The first group is mainly concerned on detecting thin objects in 2D and 3D images. The second group is a set of general approaches used in object detection and recognition.

The first group contains methods designed for thin object detection, including filtering [5], differential geometry and Hessian-analysis [6], mathematical morphology [7], and others [8]. Nevertheless, most of these approaches were designed for vessel detection, and are general tools for thin object detection.

The second group is a set of general approaches found in the object detection and recognition literature. Keypoint based methods, such as Scale-invariant Feature Transform (SIFT) [9], covariant region descriptors [10], dense descriptors [11], binary descriptors [12], and color histograms [13] are very popular approaches in object recognition literature. Although methods like SIFT and Speeded Up Robust Features (SURF) are shown to be powerful for many problems, they are patented.

2 Methods

A microfilaria is an elongated object, which can be thought as a local structure spanning a path. Therefore, a voting scheme can be used to reveal how the local structure is varied along the object. In addition, other appearance features, such as color can be used. Under this assumption, the proposed approach uses both the shape and color information for recognizing microfilariae from other coexisting objects in the 2-D microfilariae images.

2.1 Image segmentation

The input is red-green-blue (RGB) sample degraded by an illumination bias field. To make the illumination uniform, the image is first converted to grayscale and illumination is made uniform using a 2-D polynomial function [7]. Thereafter, a tophat transform is used to highlight the microfilariae. The mean-c method presented by Fisher *et al.* [14], is then applied with window of radius 75 pixels and an intercept, c, of 0.2 to obtain a binary image. The medial axis ζ is then obtained by [15]. The Meijster's approach [16] is used to calculate the Euclidean Distance Map (EDM) ψ.

2.2 Object Representation

Binary objects can be represented in different ways, such as Quench function [17] which associates the pair $\{p, \psi(p)\}$, where p is a point in the medial axis of the object and $\psi(p)$ is the EDM at p. This way an object is represented as a vector of these pairs. In this work, this concept is used to create a feature vector for the object shape. A histogram $\mathbf{h}$ of an object P is created such that, $\forall p \in \zeta(P)$, the magnitude at bin k is:

$$h_k(P) = \sum \left[\psi(p) = k\right], \quad k = 1, \cdots, \max\left(\psi(p)\right) \qquad (1)$$

where $[\square]$ is the Iverson bracket, *i.e.* $[\square] = 1$ if $\square$ is true, and 0 otherwise. This representation is similar to the generalized cylinder descriptor [18]. In (1), an object is represented by how many maximal disks it has and of what sizes. The spatial information is lost in this kind of representation; nevertheless, for objects composed of a local structure that spans the medial axis, *e.g.* nematodes, this representation seems fine. This is because microfilaria-like objects are free-form, making spatial information difficult to use. Figure 1 shows synthetic objects and their corresponding histograms. An ideal microfilaria-like object (top row) produces almost a single response (sparse vector). $\mathbf{h}$ is sparse if most of its elements are zeros. On other words, let the support of $\mathbf{h}$ be $\rho(\mathbf{h}) = \{1 \le k \le \max(\psi(p)) | h_k \neq 0\}$, the cardinality $\text{Card}(\rho(\mathbf{h})) \ll \max(\psi(p))$ [19]. In this histogram, the "energy" is concentrated at a single bin because a microfilaria is supposed to be composed of a single local structure.

When microfilaria bends or overlaps with itself or other particles, a reasonable amount of energy will still be available in the histogram representing the microfilaria. Since noise in the contour will produce a noisy skeleton (Figure 1), the approach presented by Telea *et al.* [20] is used to prune the noisy juts from the skeleton. In Figure 2, it is shown that, skeleton pruning produces better histograms. For instance, in spite of the overlapping of the object in the third row, the histogram is still similar to the ideal histogram in the first row. This makes this represen-

tation helpful for featureless objects under non-rigid deformations.

2.2.1 Incorporating Color Features

In light microscopy, intensity and color variations affect the microfilariae and the background. Several reasons contribute to these variations. For instance, the models of the imaging devices may result in noticeable variation. Color characteristics, filters, and intensity and concentration (and acidity) of the dye are other important reasons for color variations. Besides pure color variations, there are also variations in luminance.

Since the intensity variations are encoded in the color space channels along with the pure color variations, they become more difficult to tackle. However, color space transformations may help to cope with many of these variations. Separation between intensity and color information may be of a major importance.

Current literature does not make it clear which is the best color space for microfilariae, therefore, the hue-saturation-value (HSV) color space was empirically chosen in this work. Using our dataset, Figure 3 compares the color histogram descriptors [13] of the common color spaces RGB, HSV, red-green (rg), and cyan-magenta-yellow-black (CMYK), and some of the Commission Internationale de l'Eclairage (CIE) color spaces, such as CIE Lightness-ab (CIELAB), CIE Lightness-uv (CIELUV), and the tristimulus values (XYZ). In this figure, it is clear that, HSV color space overcomes the others.

2.2.2 Binary Classification

In this work, microfilariae recognition is handled as a binary classification problem. Let $\mathbf{h}$ be the shape feature vector calculated by (1) and, $\mathbf{u} = [u_{\mathrm{H}} \quad u_{\mathrm{S}} \quad u_{\mathrm{V}}]$ be the color feature vector, where $u_{\mathrm{H}}, u_{\mathrm{S}}$, and u_{V} are, the color histograms of the H, S, and V channels of the HSV color space, respectively. Each of these channel histograms is normalized by the ℓ_1-norm and then, after the vectors of the channels are combined, $\mathbf{u}$ is also normalized by the ℓ_1-norm. A common practice in multi-feature classification is to concatenate different feature vectors to form a single vector. However, since $\mathbf{h}$ is sparse and $\mathbf{u}$ is roughly dense, it is impractical to concatenate them. For this reason, the results of the classifiers were combined.

The margin Support Vector Machines classifier (C-SVM) [21] was used for both $\mathbf{h}$ and $\mathbf{u}$. Given n d-dimensional training vectors $\mathbf{v_i} \in \mathbb{R}^d$, $i = 1, \cdots, n$, and

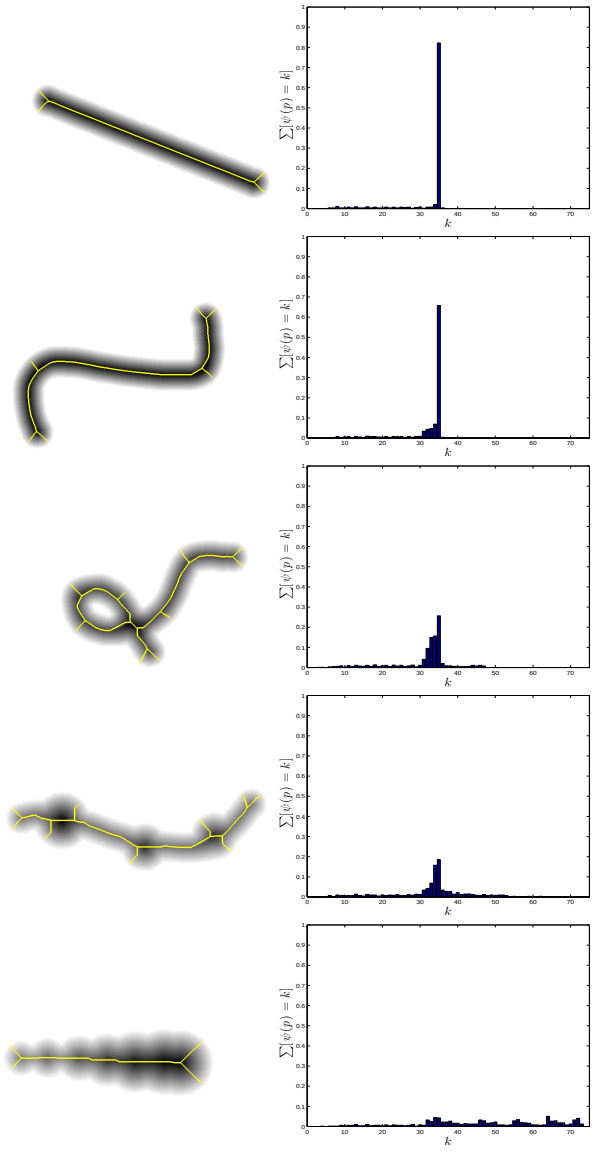

Figure 1: Histogram representation of synthetic objects without skeleton pruning. Left the medial axis superimposed on the object, and right, the object's histogram.

a labeling vector $\gamma \in \{1, -1\}$, the objective of C-SVM is to learn a mapping $\mathbf{v_i} \mapsto \gamma, \forall i$, by solving the following optimization problem:

$$\begin{aligned} \min_{\omega} \quad & \frac{1}{2}\omega^\top Q\omega - \mathbf{1}^\top\omega \\ \text{subject to:} \quad & \gamma^\top\omega = 0, \\ & 0 \le \omega_i \le C \end{aligned} \tag{2}$$

where, $\mathbf{1}$ is an all-one vector, and C is a regularization parameter. $Q = \gamma_i\gamma_j\mathcal{K}(\mathbf{v_i}, \mathbf{v_j})$ is a positive semidefinite matrix with $\mathcal{K}$ being a kernel function (*e.g.* a Gaussian), and $j = 1, \cdots, n$. The thresholding function will then be de-

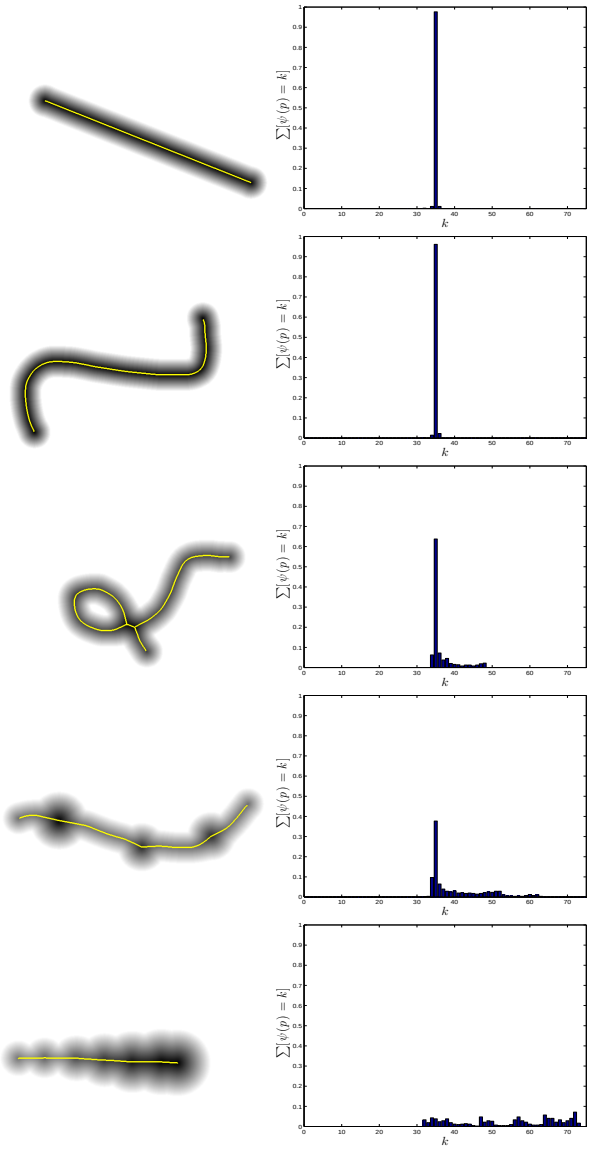

Figure 2: Histogram representation of synthetic objects with skeleton pruning. Left the medial axis superimposed on the object, and right, the object's histogram.

fined as:

$$\operatorname{sgn}\left(\sum_{i=1}^{n}\gamma_i\omega_i K(\mathbf{v_i},\mathbf{v})+b\right) \tag{3}$$

where $\mathbf{v}$ is a test vector to be classified, and b is the intercept.

Since the decision function in (3) yields crisp values, *i.e.* a label, it is difficult to establish a multi classifier ensembling strategy using such values. Therefore, rules [22] based on posterior probability estimation from SVM output [23] are used. The rule used in this work is to take the maximum between the posterior probabilities of the scores of the two classifiers.

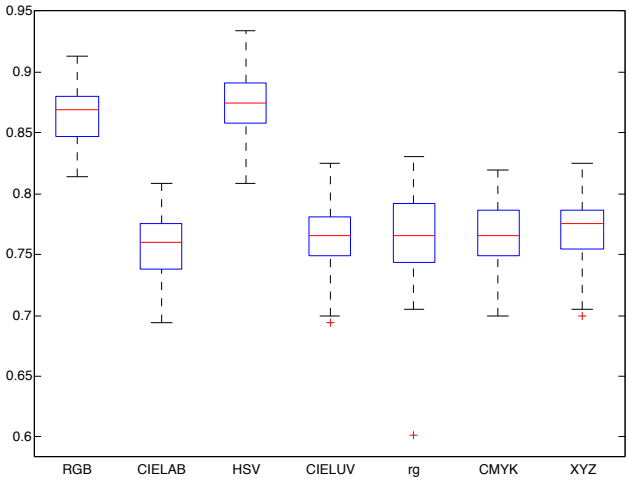

Figure 3: Color histogram [13] evaluation using several color spaces: red-green-blue (RGB), CIE Lightness-ab (CIELAB), hue-saturation-value (HSV), CIE Lightness-uv (CIELUV), red-green (rg), cyan-magenta-yellow-black (CMYK), and tristimulus values (XYZ).

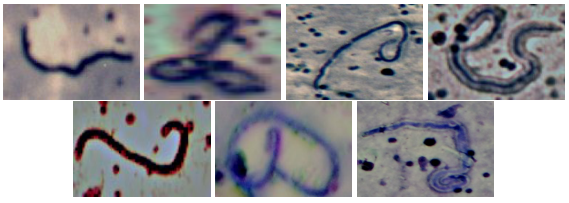

Figure 4: Positive sample patches.

3 Results

The material used in this work are randomly collected samples from two sets of image slices (produced at the Institut de Recherche pour le Développement (IRD) Montpellier - France). The first set is a collection of 32-bit JPEG RGB images of dimension 1424 × 2144 pixels. These were collected from a Zeiss Scope A1 Axio microscope with a Nikon D90 camera mounted on it, with ISO set to 400. The second set is a collection of 32-bit BMP RGB images of dimension 1536 × 2048 pixels, collected from unknown devices. The zoom factors were of 4x, 5x, and 10x.

The evaluation set is a dataset of 610 samples (patches), where each image contains only one object. These patches were manually cropped from the main images and assigned to two labels: microfilaria (positive – Figure 4) or non-microfilaria (negative – Figure 5). The numbers of positive and negative samples were of 329 (53.93%) and 281 (46.07%), respectively.

To evaluate the presented approach, it is compared with three of very successful methods in object recognition: SIFT [9] (and its recent extension Opponent-SIFT [24]), Local Intensity Order Pattern (LIOP) [10], and color

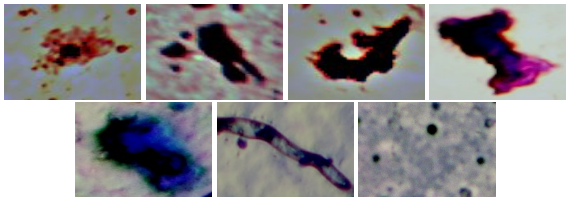

Figure 5: Negative sample patches.

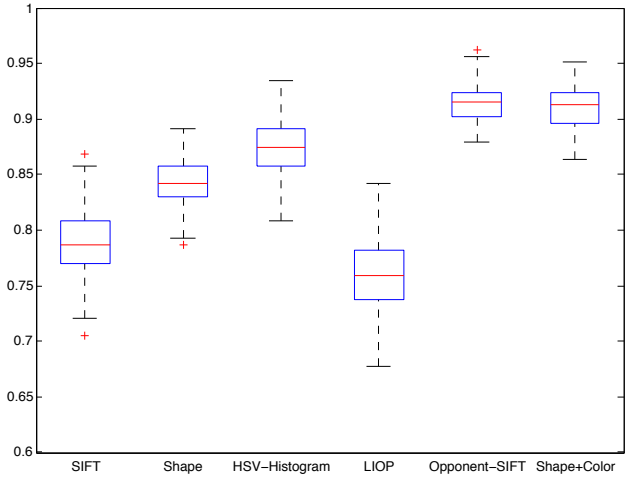

Figure 6: The accurate identifications of microfilariae for 100 runs of the descriptors: Scale-invariant Feature Transform (SIFT), shape histogram (Figure 2), Local Intensity Order Pattern (LIOP), HSV-histogram, Opponent-SIFT, and the proposed multi-classifier approach for both shape and color features.

histograms [13] using HSV color space. To that end, 100 random sub-sampling validations were performed, at each run, 70% of the dataset were chosen randomly to be used for training the classifiers, and 30% for testing. Figure 6 plots the results of each method.

4 Discussion

From Figure 6, it can be noticed that, the proposed approach outperforms the other approaches when using grayscale only (shape). However when adding color information, the proposed approach (shape + color) reported 93.5% and 87.6% of mean recall and specificity, respectively, whereas the Opponent-SIFT reported reported mean recall of 88.0% and specificity of 95.2%, respectively, which makes the proposed approach a very competitive alternative for microfilariae description given its simplicity of implementation, and the fact that SIFT is patented.

5 Conclusions

An approach for *L. loa* microfilariae recognition in 2-D images has been presented. The input sample is an RGB image containing a single object. The objective is to classify this object as microfilaria or non-microfilaria. In order to do that, image is binarized, and the skeleton and the distance map are calculated. A histogram is then calculated to represent the binary object as a function of width of the object at each skeleton point. Moreover, color features are calculated using the color histogram. Two *C*-SVM classifiers were trained on both the shape and color feature vectors and the result is fused by taking the maximum posterior probability of each classifier output. The experimental results show a good performance compared to very popular approaches, SIFT and LIOP. The proposed method is very competitive to SIFT, and since SIFT is patented, the proposed method is a very good alternative to SIFT in both grayscale or color images for detecting thin objects.

Acknowledgement: This work was conducted as part of the project "Rapid identification of high levels of Loa loa microfilaremia: a "test and treat" strategy for mass drug administration" funded by the Bill and Melinda Gates Foundation (Grant ID : OPP1033740). This work was also supported by Infectiopôle Sud, Marseilles (France) and Thamar University scholarship (Yemen).

References

[1] W.H. Organization. Bench aids for the diagnosis of filarial infections. 1997.

[2] Gardon J., Gardon-Wendel N., Ngangue D., Kamgno J., Chippaux J., Boussinesq M., Serious reactions after mass treatment of onchocerciasis with ivermectin in an area endemic for loa loa infection, The Lancet, 1997, 350(9070),18 – 22.

[3] Boussinesq M., Loiasis, Annals of tropical medicine and parasitology, 2006, 100(8), 715–731.

[4] Knott J., A method for making microfilarial surveys on day blood, Transactions of the Royal Society of Tropical Medicine and Hygiene, 1939, 33(2),191 – 196.

[5] Chaudhuri S., Chatterjee S., Katz N., Nelson M., Goldbaum M., Detection of blood vessels in retinal images using two-dimensional matched filters, IEEE Transactions on Medical Imaging, 1989, 8(3), 263–269.

[6] Frangi A.F., Niessen W.J., Vincken K.L., Viergever M.A., Multiscale vessel enhancement filtering, In International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Lecture Notes in Computer Science, Springer Berlin Heidelberg, 1998, 1496, 130–137.

[7] AL-Tam F., dos Anjos A., Bellafiore S., Shahbazkia H.R., Detection of root knot nematodes in microscopy images. Proceedings

of the International Conference on Bioimaging, (12-15 January 2015, Lisbon, Portugal), 2015, 76-81.
[8] Tankyevych O., Talbot H., Dokladal P., Curvilinear morpho-Hessian filter, Proceeding of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro. (14-17 May 2008, Paris, France), IEEE, 2008, 1011–1014, 2008, Paris, France.
[9] Lowe D.G., Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, Springer, 2004, 60(2), 91–110.
[10] Wang Z., Fan B., Wu F., Local intensity order pattern for feature description, Proceedings of the IEEE International Conference on Computer Vision (ICCV), (6-13 November 2011, Barcelona, Spain), IEEE, 2011, 603–610.
[11] Dalal N. and Triggs B., Histograms of oriented gradients for human detection, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (20-26 June 2005, San Diego, CA, USA), IEEE, 2005, 886–893.
[12] Ojala T., Pietikäinen M., Harwood D., A comparative study of texture measures with classification based on featured distributions, Pattern Recognition, Elsevier, 1996, 29(1), 51–59.
[13] Swain M.J., Ballard D.H. Color indexing, International Journal of Computer Vision, Springer, 1991, 7(1),11–32.
[14] Fisher R.B., Perkins S., Walker A., Wolfart E., Hypermedia Image Processing Reference, J. Wiley & Sons Publishing, 1996.
[15] Lam L., Lee S.W., Suen C.Y., Thinning methodologies-a comprehensive survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, 14(9), 869–885.
[16] Meijster A., Roerdink J.B., Hesselink W.H., A general algorithm for computing distance transforms in linear time, Mathematical Morphology and its applications to image and signal processing, Springer, 2002, 18, 331–340.
[17] Luc V. Dougherty E.R., Morphological segmentation for textures and particles, Digital image processing methods, CRC Press, 1994, 42, 43 – 102.
[18] Agin G.J., Binford T.O., Computer description of curved objects, IEEE Transactions on Computers, 1976, C-25(4), 439–449.
[19] Starck J.L., Murtagh F., Fadili J.M., Sparse image and signal processing: wavelets, curvelets, morphological diversity, Cambridge University Press, 2010.
[20] Telea A., Van Wijk J.J., An augmented fast marching method for computing skeletons and centerlines, Proceedings of the symposium on Data Visualisation, (27-29 May 2002, Barcelona, Spain), Eurographics Association, 2002, 251–259.
[21] Chang C.C., Lin C.J., Libsvm: A library for support vector machines, ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3), 1-26.
[22] Kuncheva L.I., Combining pattern classifiers, Methods and Algorithms, Wiley, Chichester, 2004.
[23] Platt J., Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, Advances in large margin classifiers, 1999, 10(3), 61–74.
[24] van de Sande K.E., Gevers T., Snoek C.G., Evaluating color descriptors for object and scene recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9),1582–1596.